RULE LEARNING IN EXPERT SYSTEMS USING GENETIC ALGORITHM: 2, EMPIRICAL STUDIES

Kwong Sak Leung[+], Yee Leung[++], Leo So[+++], and Kin Fai Yam[+]

[+]Department of Computer Science
[++]Department of Geography, and
Center for Environmental Studies
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
[+++]Computer Science Department
University of Arizona
Tucson, AZ 85721, U.S.A.

Keywords: Genetic algorithm, rule learning, knowledge engineering, token competition, rule migration, expert system.

Abstract

Token competition and rule migration are two special features of *SCION*, a platform for developing genetic algorithm (GA) based applications. Their performances in automatic rule learning are evaluated by a series of empirical studies. It is demonstrated that GAs with token competition has a clear advantage over those without it. Except for some situations, GAs with rule migration has a better performance on the average.

1. Introduction

In part 1 of this two-part series of paper, basic concepts of *SCION* (a platform for developing GA based applications) are proposed [1]. It is argued that token competition and rule migration are important mechanisms which can improve the efficience of automatic rule learning. The purpose of this part of the series is to demonstrate with empirical studies the effectiveness of the proposed mechanisms.

Case studies and empirical results using token competition are examined in section 2. Results of a series of empirical studies using rule migration are analyzed in section 3. Comparison of *SCION* and a conventional GA system is then made in section 4. The paper is concluded with some directions for further research.

2. Token Competition — Empirical Results and Interpretations

In order to evaluate the effectiveness of GAs with token competition, two empirical studies are carried out as follows:

(a) Problem with equal clusters. For each class, the data points are gathered to form four clusters with compatible size.

(b) Problem with unequal clusters. For each class, the data points are gathered to form four clusters with one large (in terms of area covered) cluster.

It can be observed in Figs. 1 and 2 that *SCION* with token competition significantly outperforms the one without it. In both cases, the one with no competition get stuck after certain generations since most of the rules in the population look like each other and no diversity is created even after the crossover operation has taken place.

Roughly speaking, about 50% of the parent rules should be disposed so that more room is spared to accommodate a larger set of children rules. Therefore, the diversity of search is greatly enhanced without expanding the total population.

Towards the end, the number of redundant rules killed remains constant. This further confirms the importance of token competition. After the population has evolved for a certain period of time, only a few viable genes (i.e. particular duples) remain in the population. Therefore, the chance of producing a redundant child rule is very high.

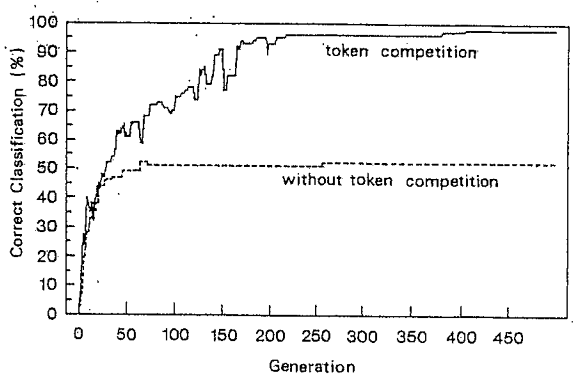


Fig. 1 Performance with and without token competition in four-equal-cluster problem

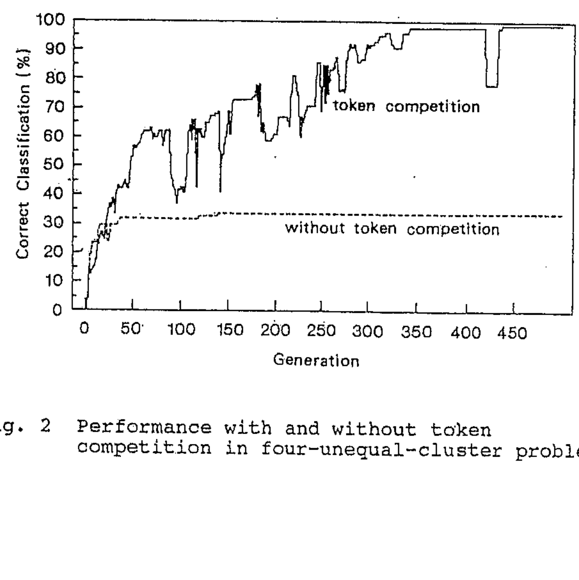


Fig. 2 Performance with and without token competition in four-unequal-cluster problem

3. Rule Migration — Empirical Results and Interpretations

The objective of Rule Migration (RM) is clear: to preserve the evolution effort so as to shorten the evolution time. Since Rule Migration is a new concept in GAs, its capability is still uncertain. So experiments are conducted to examine the:

(a) Performance of RM with different problem complexities,
(b) Performance of RM with different migration quotas,
(c) Performance of RM with different population sizes,
(d) Performance of RM with different volumes of training data,
(e) Robustness of RM with different initial rule sets.

3.1 Different Problem Complexities

In this section, the capacity of migration is tested in different problems with different levels of complexities. Two sets of experiments are carried out. Each set consists of three problems: (a) an easy two-dimensional problem (2 attributes) with 2 different classes each, (b) a three-dimensional problem (3 attributes) with 3 different classes and (c) a difficult four-dimensional problem (4 attributes) with 4 different classes. Problems In set #2 are more complex than set #1 because their training data scatter more and are grouped into four different clusters.

In easy problems, GAs with and without migration give the same results. The total generations required are too little that migration has no advantage. Evolution without migration is already adequate. However, in more complex and difficult problems like set #2, migration improves the performances. Since the data for each class scatter a lot, most of the rules in a particular class would have a large number of false alarms to other classes and become bad rules themselves, resulting in having little survivors in each class. Without migration, the progress made in the first few generations is very limited because of bad parents. However, since the data scatter, the bad rules for a particular class have a very high probability as a good one in other classes.

Table 1. Set #1: 1 cluster/class

Class #	Attribute #	Generation Number for 100% learning: Without Migration	With Migration (% of w/o Migration)
2	2	3	3 (100%)
3	3	3	3 (100%)
4	4	3	3 (100%)

Table 2. Set #2: 4 clusters/class

Class #	Attribute #	Generation Number for 100% learning: Without Migration	With Migration (% of w/o Migration)
2	2	98	52 (53.6%)
3	3	95	78 (82.1%)
4	4	289	155 (53.6%)

Rule migration can thus inject vitality to the system at the beginning of evolution and set a good path for further evolution. We can see in Tables 1 and 2 that migration always performs better in the first few generations and set a better evolution direction.

3.2 Different Migration Quotas

The role of quota in the performance of rule migration is very subtle. Without rule migration, it usually takes more time for rules to evolve. If it is set too large, it will hinder the evolution, because the more foreign immigrants a population has, the less the offsprings the original rules can produce. It can be shown that most rule immigration happens in the first few generations of evolution. (See disscussion on the pattern of rule migration in part 1 of this series.) As the first few generations is very critical for the system to establish its own evolution path, different quota limits exert different effects on the system performance.

Two sets of experiments are carried out, the former one uses 100 training data for each class while the latter uses 400. The results are obtained in Table 3.

These two sets have similar results. Evolution time decreases from quota 1 to a minimum in quota 2. Then, it rebounds afterward and levels off at larger quotas.

It is because in each generation there are only a handful of qualified immigrants. With insufficient quota, some useful immigrants would be blocked. Also, when the quota exceeds the number of qualified immigrants, pushing up the quota limit would not have any effect.

Table 3

Migration Quota	Generation Number for 100% learning: 100 Training Data	400 Training Data
1	135	79
2	111	51
3	123	78
4	190	78
5	190	78

3.3 Different Rule Population Sizes

In this experiment, classification rules are learnt using the same training data. The initial population sizes of the rule sets are 10, 20, 40 and 80 in each class respectively. The time for *SCION* to learn all the data is obained in Table 4.

Table 4

Rule Population	Generation Number for 100% learning: Without Migration	With Migration (% of w/o Migration)
10	more than 2659	1215 (45.7%)
20	298	517 (173.5%)
40	95	78 (82.1%)
80	29	87 (300.0%)

From the results we notice that migration is more powerful in small population. In the case with 10 rules, the room for producing offsprings is very limited (about 5 only), and the evolution rate is very low. We can see that with no migration, using 10 rules still cannot reach 100% after 2659 generations. However, with migration, one competent rule from another class can immediately inject vigour into the system and it requires only 1215 generations to reach 100%. So for a relatively small population, rule migration should be employed.

As the population size increases, the number of offsprings also increases and the significance of rule migration decreases drastically and behaves worse than having no migration at all.

3.4 Different Training Data Volumes

In this experiment, classification rules are learnt using the same initial rule set and same rule population size of 40. However, the total number of training data for each class vary from 100 to 800. The results for learning all the data are tabulated in Table 5.

Table 5	Generation Number for 100% learning	
Training Data #	Without Migration	With Migration (% of w/o Migration)
100	62	123 (198.4%)
200	152	108 (71.1%)
400	95	78 (82.1%)
800	136	82 (60.3%)

Again migration performs better in cases with large number of training data. As the number of training data increases, the data will be more diverse but more representative. Therefore, the difficulty of classifying all of them correctly increases as the volume of training data increases. Migration, on the other hand, can overcome this difficulty by setting new directions guided by the immigrants. So for real applications where there are thousands of training data in each class, it is better to use migration.

3.5 Different Initial Rule Sets

The objective of this experiment is to test whether migration can always do better for the same problem, i.e. with the same number of classes and attributes, and with the same rule population sizes. To test this, different initial rule sets are used for learning with and without migration. The results are depicted in Table 6.

It can be seen that having migration has a better performance than not having migration on the average. This confirms our belief that migration can set an efficient evolution path.

Table 6	Generation Number for 100% learning	
Initial Rule Set	Without Migration	With Migration (% of w/o Migration)
#1	95	78 (82.1%)
#2	60	57 (95.0%)
#3	69	75 (108.7%)
#4	197	71 (36%)
Average:	105.25	70.25 (66.7%)

3.6 General Pattern of Rule Migration

An experiment is done to study the pattern of migration in terms of the number of immigrants and the minimum requirement for immigration to a particular class at each generation. The results are depicted in Fig. 3.

We can see that the minimum requirement for immigration increases while the number of qualified immigrants decreases as the evolution goes.

At the beginning of an evolution, since the rules in each class are not so strong, the minimum requirement for entering a class is relatively low. Therefore, the bad rules eliminated from a class can easily go to other classes.

As the rules in each class become stronger, the threshold for immigration to a foreign class becomes higher. So, the number of immigrants decreases as the difficulty of immigration increases.

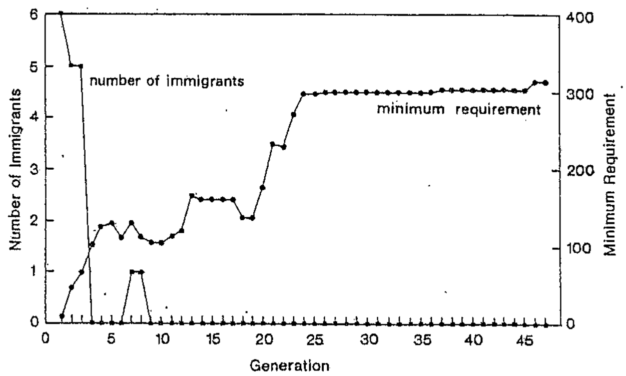


Fig. 3 Pattern of rule migration

3.7 A Note on Rule Migration

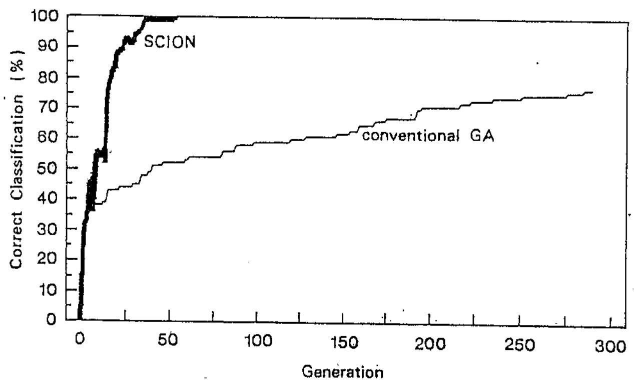
Although, the average performance of GA with rule migration is better than the one without it, it should be emphasized that rule migration is not a panacea in rule learning. It behaves well in problems with large classes, large cluster number, and large training data set. Also it performs exceptionally well when rule population is relatively small comparing to the problem itself (e.g. rule population of 10 in problem with 5 classes with 5 clusters each, 10 attributes and 2000 training data for each class).

However, migration should be employed together with the correct migration quota to achieve its maximum benefits. So it is hoped that the later version of *SCION* can dynamically adjust the migration quota or even the usage of migration for different problems.

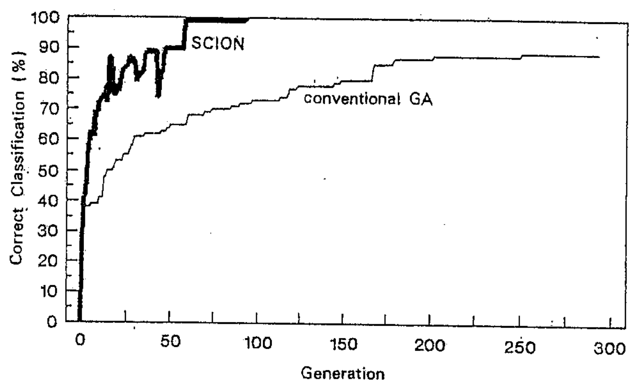
4. *SCION* vs a Conventional GA System

To demonstrate the superiority of *SCION*, its performances is compared with that of a GA system built with conventional features [2]. The results are depicted in Figs. 4a, b.

In clustered problem space, *SCION* easily outperforms the traditional GA by 10 to 15 times, with even a smaller population. But we should not ignore that the traditional tree representation, which allows simple linear expression, may perform better with data embedded with some mathematical properties. Also, since only the strongest rule is supposed to classify all the identical data and the evaluation of a rule strength is biased by the size of the rule, GA may be able to give a set of more concise production rules at the end.



(a) *SCION*: 40 rules Conventional GA: 64 rules
3 classes: 4 equal clusters per class



(b) *SCION*: 40 rules Conventional GA: 64 rules
3 classes: 1 big cluster out of 4 per class

Fig. 4 *SCION* vs conventional GA

5. Conclusion

In designing *SCION* many concepts of evolution such as selection of parents, candidates for mutation, punishment and award are constructed from the perspective of eugenics and social planning. Also we can observe in our GA system many interesting phenomena in human society, such as the negative impact of inbreed crossover and the propagation of good genes down to the offsprings. We are convinced that GA is a promising tool to tackle many optimization problems and its ability to get rid of local maxima is quite impressive. As it does not depend on any prior knowledge about the problems, GA is an ideal general learning mechanism in rule-based systems. *SCION* is going to be further enhanced to take into consideration of fuzzy knowledge [3], training and synthesis of neural networks [4].

References

[1] K.S. Leung, Y. Leung, Leo So, and Kin Fai Yam," Rule Learning in Expert Systems using Genetic Algorithm: 1, Concepts," Proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks, IizukA, Japan, July 1992.

[2] Richard Forsyth, "Beagle — a Darwinian approach to pattern recognition," Kybernetes, Vol. 10, pp. 159-166, 1981.

[3] Chuck Karr, "Genetic Algorithms for Fuzzy Controllers," AI EXPERT, pp. 26-33, February 1991.

[4] Maureen Caudill, "Evolutionary Neural Networks," AI EXPERT, PP. 28-33, March 1991.